

Building A Scientific Data Analysis Server

Deb Agarwal

Berkeley Water Center, University of California, Berkeley

<http://bwc.berkeley.edu>

Collaboration with Microsoft Research



Excellent Data Archives Available

AmeriFlux



About AmeriFlux
[Objectives](#)
[Organization](#)
[Strategic Plan](#)
[Participants/Sites](#)
[Participant Information](#)
[Site Information](#)
[List Server](#)
[Standards](#)
[Core/Desired Measures](#)
[Guidelines/SOPs](#)
[Data](#)
[Submission Guidelines](#)
[Available Data](#)
[Data Access](#)
[Key Resources](#)
[Equipment](#)
[Sources/Tips](#)
[Publication Lists](#)

AmeriFlux
 The AmeriFlux to a workshop CO₂ and Water Learn more at questions tha
 The AmeriFlux ecosystem lev momentum sp interannual ti from North An Please click o sites included
 AmeriFlux is p (FLUXNET) wh observations l more about E networks.
 If you are inte

The US Long Term Ecological Research Network

LTERR Home | Intranet

LTERR Site Home Pages Go Search

About LTERR

LTERR Network

What's New

- Moore makes 'Nature' special
- Harvard Forest and University of Massachusetts Scientists develop new process for synthesizing ecological data sets
- Jornada Researchers Helping Unravel Causes of Desertification in the Southwestern United States
- Bonanza Creek Featured in USA Today Article on Climate Change
- Cedar Creek Study Shows That Plant Diversity Enhances Ecosystem Resilience

Read More

network is a collaborative effort is investigating ecological scales. live research across sites and and international research

The LTERR program in 1980 to omena in the United States. ns and research emphases cation, network publications, and

CUAHSI

The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)



0 600 1,200 Miles

About Map Legend

Hydrology Data Access System

HDAS provides uniform access to several national-scale water resource databases:

- USGS NWIS
- AmeriFlux Network

How to use it ?

To chart or download time series, please follow these steps:

- Zoom in on area of interest until you see station points on the map
- Click on the icon

National Weather Service Hydrologic Information Center

Home News Organization

Current Flooding
 -WFO Products
 -River Locations in Flood
 -Flood Summary

Outlooks
 -Hydrologic
 -Meteorological

Current Hydrologic Conditions
 -WFO Products
 -River Conditions
 -Drought

Archive & Historic Information



3613 Total Gauges
 5 Locations in Flood

296 Gauges: Observations older than 12 hours
 17 Gauges: Out of Service

Last map update: Sun, Sep, 17, 2006 at 05:55:13 pm

Select a satellite/collection:

☒ Terra, Collection 4
☐ Aqua, Collection 4
☐ Aqua + Terra, Collection 4
☐ Terra, Collection 5
☐ Aqua + Terra, Collection 5

Last - Input the date:
 Start date: 2002/1
 End date: 2006/36
 (Format: YYYYDDDD)

Select a product:

Daily:

Top-of-the-atmosphere Radiance: ☐ MOD02/MYD02

Surface Reflectance: ☐ MOD09/MYD09 ☒ MOD09CMG/MYD09CM

Snow Covers: ☐ MOD10_L2/MYD10_L2

Surface Temperature: ☐ MOD11/MYD11

Active Fires: ☐ MOD14/MYD14

Active Fire on Surface Reflectance: ☐ MOD14MOD09/MYD14MYD09

LAI/FPAR: ☐ MOD15A1/MYD15A1

Sea Ices: ☐ MOD29/MYD29 ☐ MOD29E10/MYD29E10

8-days:

Surface Reflectance: ☐ MOD09A1/MYD09A1

Snow Cover: ☐ MOD10C2/MYD10C2

Surface Temperature: ☐ MOD11A2/MYD11A2

Fire: ☐ MOD14A2/MYD14A2

Climate Data Interim

Search Field: Search NCDC

of Climate Data
 matic
 ter
 ling the Future

Weather/Climate, Events, Information & Assessments

Satellite

Upper Air

Marine

Paleoclimatology

Based

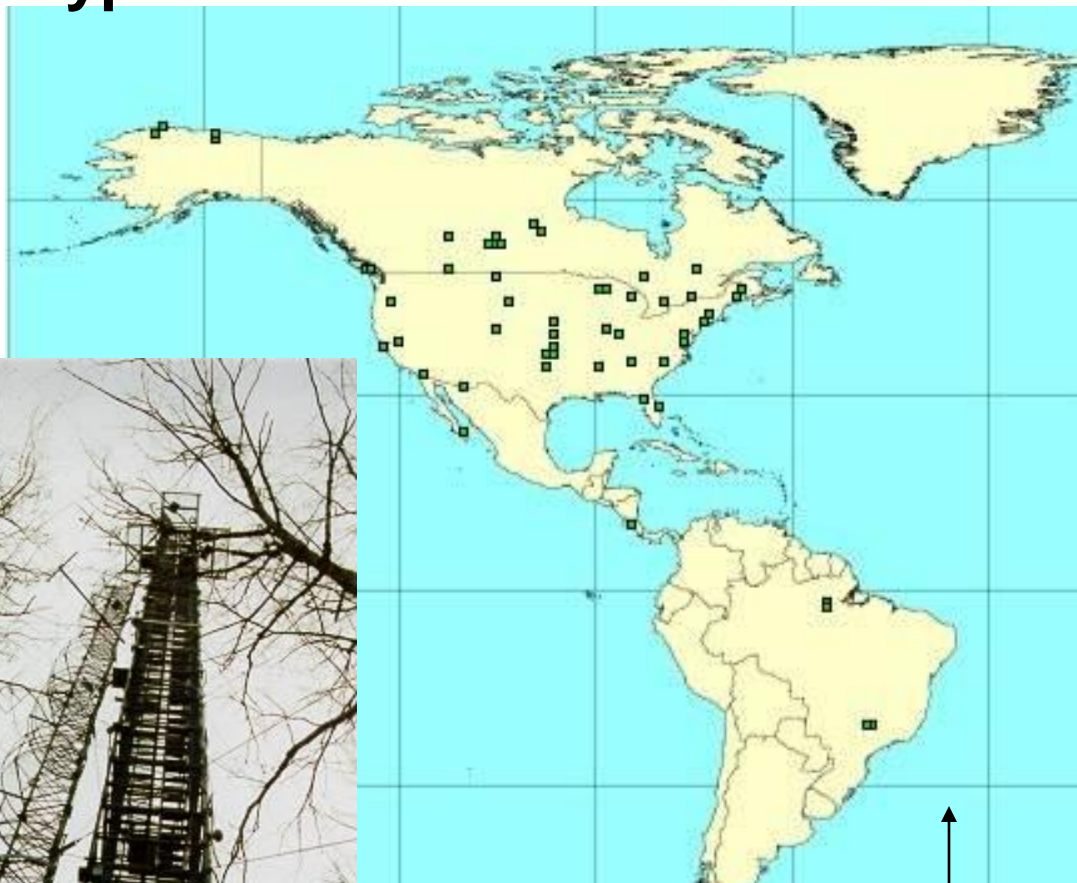
HOW ARE WE DOING?
 A user survey

FIRSTGOV
 The U.S. Government's Official Web Portal

Disclaimer

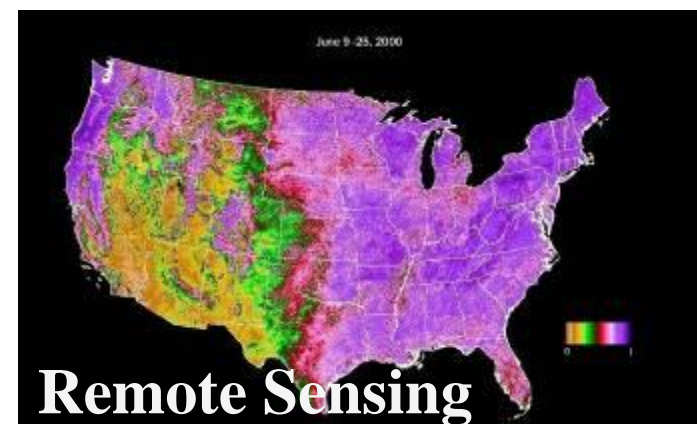
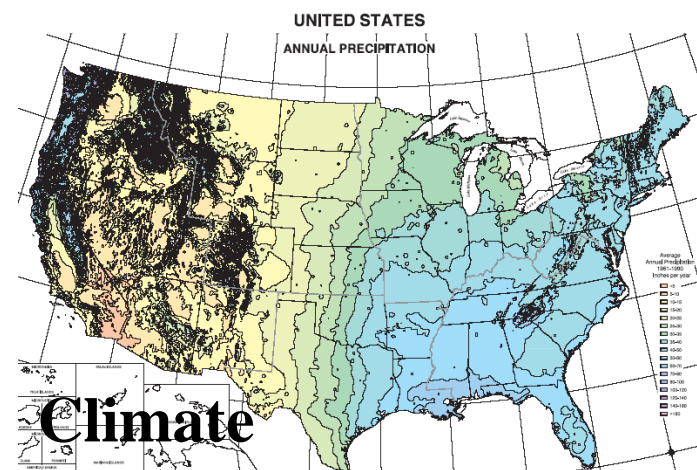
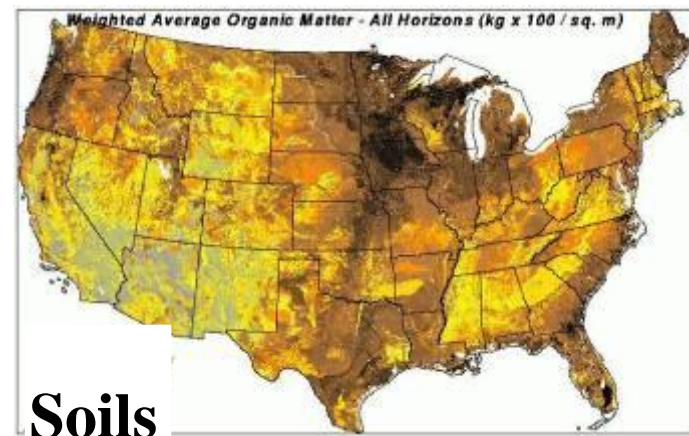
/ncdc.html
 06:18:59:37 EDT
 06:18:59:37 EDT by ncdc.webmaster@noaa.gov

Typical Carbon-Climate Datasets



Observatory
datasets

Spatially
continuous
datasets



FLUXNET Synthesis Dataset Overview

- ▶ 240 Sites around the world and 921 site years of data
- ▶ Each site reports a minimum of 22 and more typically 35 common measurements.
- ▶ Total data reported to date on the order of 800M half-hourly measurements.
- ▶ Communal science – each principle investigator acts independently to prepare and publish data.
- ▶ Second level data published to and archived at regional networks.
- ▶ Uniform quality, gap filling, and science variable derivation to generate higher level data sets.

<http://www.fluxdata.org>



FLUXNET Synthesis Support Server

The screenshot displays the FLUXNET Synthesis Support Server interface. The top navigation bar includes links for Home, Dataset Information, Blog, User Support, Data Download, and Data Browsing. The main content area is titled "Dataset Information" and provides an overview of the dataset and participating sites. A sidebar on the left lists various site categories and data availability information.

The "CumulativeValuesByVegType" section is highlighted, showing an Excel Web Access view of the dataset. The Excel spreadsheet displays cumulative annual variable values by site, with columns for site information, variables (Precip, Rd, Rq, Rn, SWC, Ta, Ts), and dates. The data is organized into a PivotTable structure, with a PivotTable Field List visible on the right side of the spreadsheet.

Dataset Information

Fluxnet-LaThuile > Dataset Information

Information About Participating Sites and Available Data at Those Sites

The links below provide an overview of the dataset and the sites that contributed data to the dataset. They also provide some specific information about the datum types and sites in this dataset as well as data

Maps and Imagery

- Map of Site locations
- Site imagery

Site Ancillary and Available Data

- Summary of key site ancillary data
- General Site Ancillary Data
 - Koeppen-Geiger climate classification
 - IGBP vegetation classification
 - Full Site Ancillary Data
- List of Sites by Country
- List of Sites by Climate
- List of Site Curators
- List of Sites by Network

Data Availability Information

- By Site List of Yearly Data
- By Site List of Cumulative Data

Data Explanations (in Excel)

- List of variables in the dataset
- List of variables in the dataset
- Datumtype report
- Datumtype/Offset
- Equations used in the dataset
- List of calculation

List of the Proposal Table

CumulativeValuesByVegType

Excel Web Access - CumulativeValuesByVegType

Open | Update | Find

Cumulative Annual Variable Values by Site

Download Excel

Working Group: All Current Fluxnet Daily

Site	Precip	Rd	Rq	Rn	SWC	Ta	Ts
mm y-1	GJ m-2 y-1	GJ m-2 y-1	GJ m-2 y-1	m-3 m-3	°C	°C	
AT-Neu Austria - Neustift/Stubai Valley	1,284.3359		4.2914	1.3867	31.8751	6.7856	8.4992
AU-How Australia - Howard Springs	1,630.2953		6.7924	4.0136	10.3777	25.8592	30.0698
AU-Tum Australia - Tumbarumba	1,123.3560		5.5826	2.8324		9.3362	10.4106

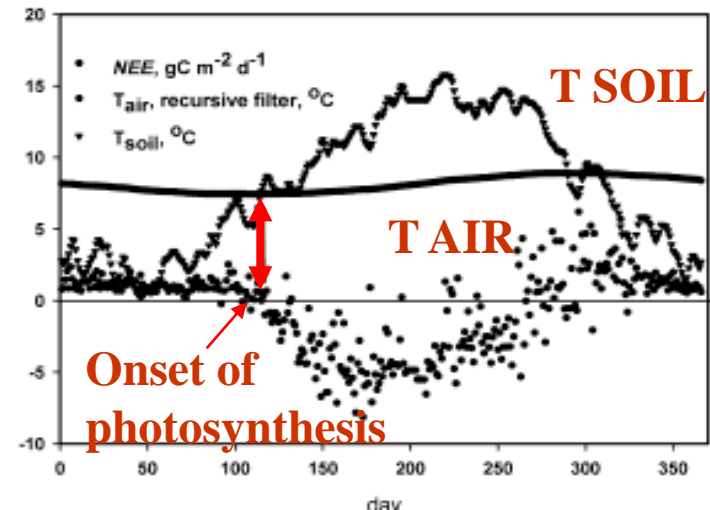
PivotTable Field List

Drag items to the PivotTable report

- ☐ Dataset
- ☐ Datumtype
- ☐ Name
- ☐ Exdatumtype
- ☐ Value units
- ☐ Repeat
- ☐ Ecosystem
- ☐ Latitude
- ☐ Site
- ☐ Day of Year
- ☐ Hour of Day

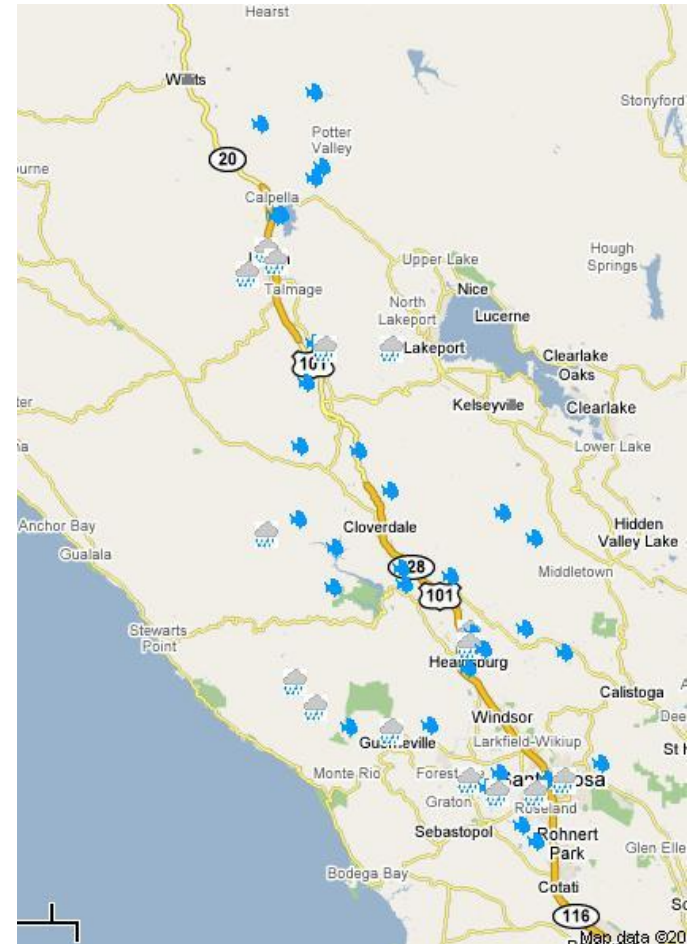
Example Carbon–Climate Investigations

- ▶ Effects of early spring on carbon uptake
- ▶ Role of vegetation, climate, and latitude on carbon flux
- ▶ Effect of disturbances on atmospheric carbon
- ▶ Linkages to remote sensing and other data sets just beginning



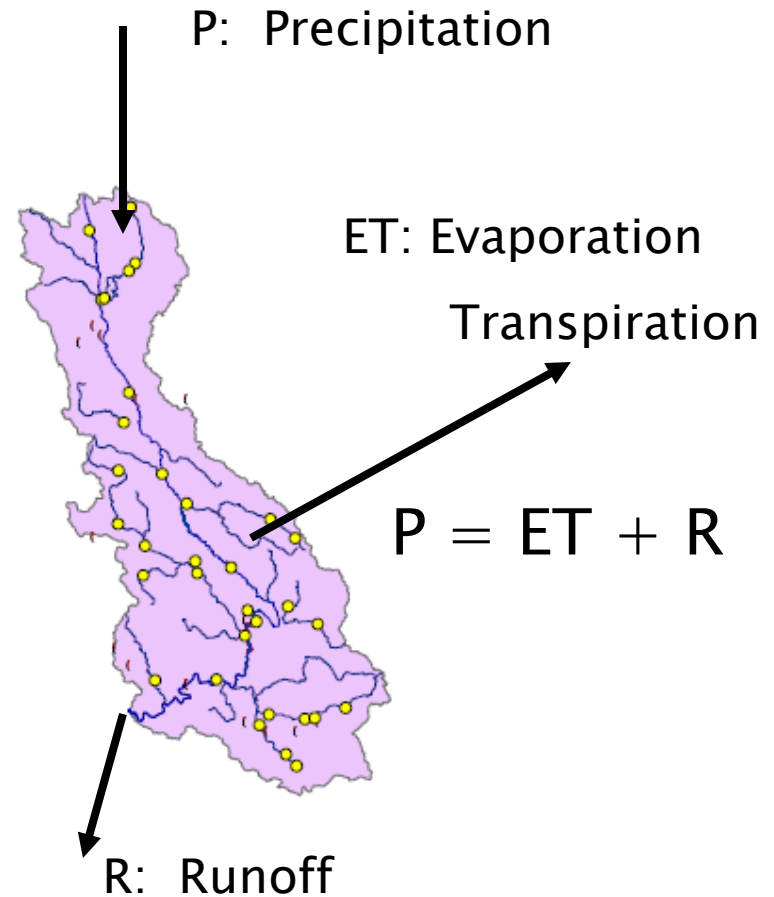
Russian River Data

- ▶ USGS Stream Gages
- ▶ NOAA Weather Stations
- ▶ County water agency
- ▶ Driller logs
- ▶ EPA gages
- ▶ Research team measurements
- ▶ Data collected to date
~23M data points
<http://bwc.berkeley.edu>



Example Russian River Investigations

- ▶ Annual water balance model for the watershed
- ▶ Historical turbidity and correlation to fish population



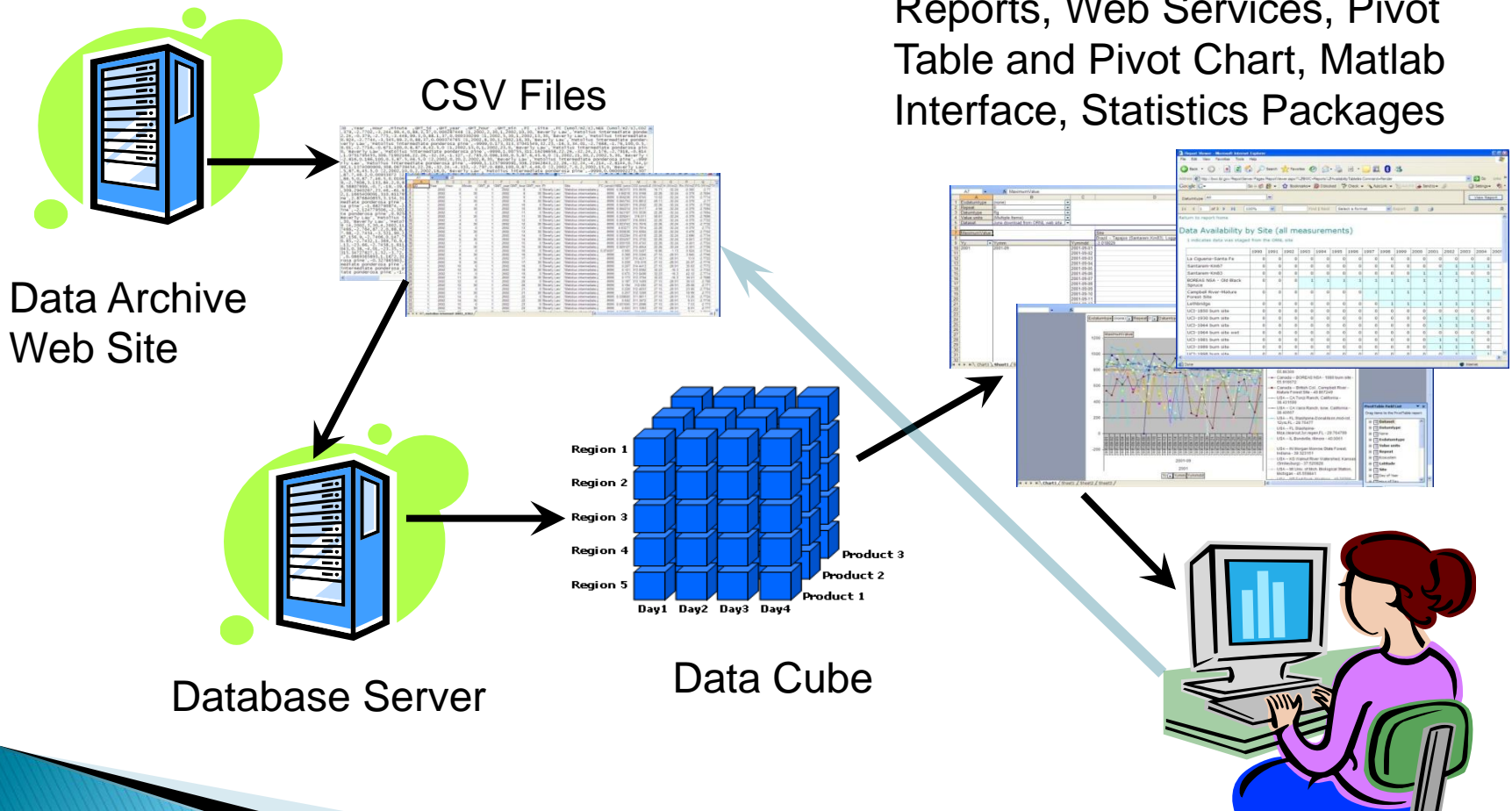
Measurements Are Often Not Simple or Complete

- ▶ Gaps in the data
 - Quiet nights
 - Bird poop
 - High winds
 - Batteries die
 - Curious hikers/animals
 - Rivers flood
- ▶ Difficult to make measurements
 - Leaf area index
 - Wood respiration
 - Soil respiration
- ▶ Localized measurements – tower footprint
- ▶ Local investigator knowledge important
- ▶ Science goals and standards are often not uniform across the groups taking measurements

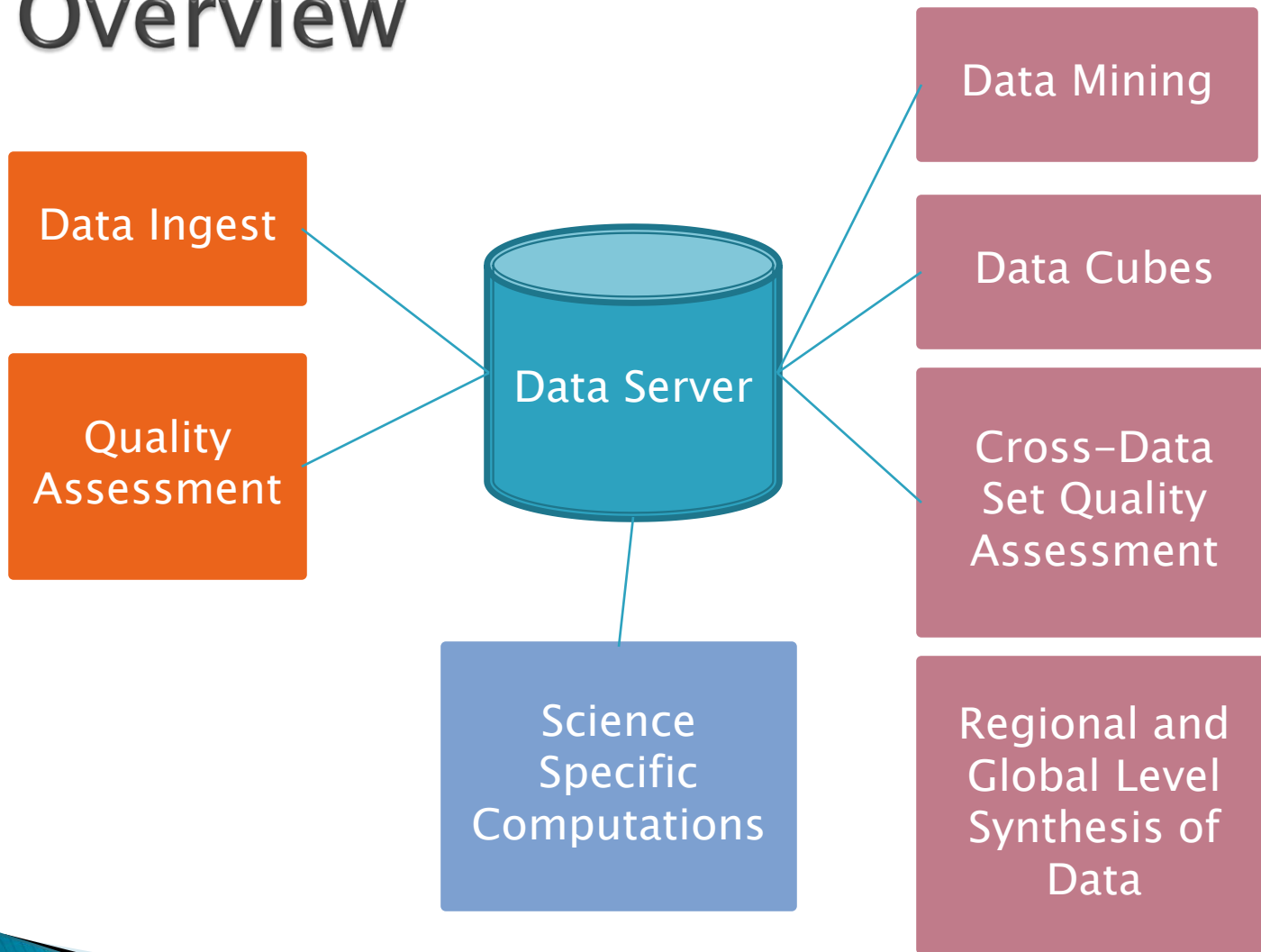


Scientific Data Server

Reports, Web Services, Pivot Table and Pivot Chart, Matlab Interface, Statistics Packages



Scientific Analysis Server – Logical Overview



Ideal Science Server

